

## KLASIFIKASI PELANGGAN PADA *CUSTOMER CHURN* *PREDICTION MODELS* MENGGUNAKAN *DECISION TREE*

### *CUSTOMER CLASSIFICATION IN CUSTOMER CHURN* *PREDICTION MODELS USING DECISION TREE*

Frans Sinata, I1834@lecturer.ubm.ac.id<sup>1)\*</sup>, Angelina Pramana Thenata,  
Angelina.pramana@outlook.com<sup>2)</sup>, Agustinus Fritz Wijaya,  
agustinus.wijaya@bundamulia.ac.id<sup>3)</sup>, I Gusti Ngurah Suryantara,  
gusti@bundamulia.ac.id<sup>4)</sup>, Jusia Amanda Ginting, jginting@bundamulia.ac.id<sup>5)</sup>, Destriana  
Widyaningrum, I0894@lecturer.ubm.ac.id<sup>6)</sup>, Ester Lumba, I0178@lecturer.ubm.ac.id<sup>7)</sup>

<sup>1)</sup> Program Studi Informatika/Fakultas Teknologi dan Desain, Universitas Bunda Mulia

Diterima 15 Juli 2025 / Disetujui 30 September 2025

#### ABSTRACT

*Intensified competition in the modern business world necessitates that companies implement proactive customer retention strategies, making customer churn prediction a paramount focus. This study aims to develop and evaluate a classification model utilizing historical customer data to proactively identify customers at risk of churn or those likely to remain loyal. The methodology employed is Data Mining, specifically the Classification technique, selecting the C4.5 Decision Tree algorithm for its strength in generating transparent and easily interpretable decision rules. The analyzed dataset comprises 996 customer samples, including vital attributes such as gender, age, payment method, and transaction history. The classification predicts customer status into one of two target categories: loyal or churn. The results indicate that the constructed model successfully classified 636 customers (approximately 63.8%) as loyal and 360 customers (approximately 36.2%) as churn, achieving a model accuracy of 98%. This finding demonstrates the C4.5 Decision Tree's effectiveness in mapping customer loyalty patterns. Practically, the model provides quantifiable insights for companies to formulate more targeted marketing and customer retention initiatives.*

**Keywords:** *Customer Churn Prediction, Decision Tree, Algorithm C4.5*

#### ABSTRAK

Persaingan yang semakin ketat dalam dunia perdagangan modern menuntut perusahaan untuk menerapkan strategi retensi pelanggan yang proaktif, menjadikan prediksi *customer churn* (pelanggan yang berhenti menggunakan layanan) sebagai fokus utama. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model klasifikasi yang memanfaatkan data pelanggan historis guna mengidentifikasi secara dini pelanggan yang berpotensi *churn* atau tetap loyal. Metode yang digunakan adalah Data Mining, khususnya teknik Klasifikasi, dengan memilih algoritma Decision Tree C4.5 karena keunggulannya dalam menghasilkan aturan keputusan yang transparan dan mudah diinterpretasikan. Dataset yang dianalisis melibatkan 996 sampel pelanggan, mencakup berbagai atribut penting seperti jenis kelamin, usia, metode pembayaran, dan riwayat transaksi. Klasifikasi dilakukan untuk memprediksi status pelanggan ke dalam salah satu dari dua kategori target: loyal atau *churn*. Hasil pengujian menunjukkan bahwa model yang dibangun mampu mengklasifikasikan 636 pelanggan (sekitar 63.8%) sebagai kategori loyal dan 360 pelanggan (sekitar 36.2%) sebagai kategori *churn*, dengan mencapai tingkat akurasi model sebesar 98%. Temuan ini menunjukkan efektivitas Decision Tree C4.5 dalam memetakan pola loyalitas pelanggan. Secara praktis, model ini berkontribusi dalam menyediakan wawasan yang terukur bagi perusahaan untuk merumuskan inisiatif pemasaran dan retensi yang lebih tepat sasaran.

**Kata Kunci:** *Customer Churn Prediction, Decision Tree, Algoritma C4.5*

\*Korespondensi Penulis:

E-mail: I1834@lecturer.ubm.ac.id

## PENDAHULUAN

**Persaingan** dalam dunia perdagangan terus meningkat secara signifikan, didorong oleh perubahan lingkungan bisnis yang cepat dan dinamis [1].[2][3] Dalam konteks ini, **loyalitas pelanggan** menjadi aset strategis utama dalam strategi pemasaran. Loyalitas yang kuat adalah faktor penting untuk mendorong **pembelian berulang**, **meningkatkan performa bisnis**, serta memiliki korelasi positif terhadap **nilai bisnis** dan kemampuan perusahaan dalam menarik pelanggan baru[4].[5]

Perusahaan saat ini menyimpan data pelanggan dalam jumlah besar (*big data*). Data mentah ini seringkali belum dimanfaatkan secara optimal untuk memahami pola perilaku krusial pelanggan. Untuk mengatasi tantangan ini, **Data Mining** muncul sebagai proses yang mampu **mengekstraksi informasi berharga** dari data historis.

Secara spesifik, teknik data mining dapat digunakan untuk **memprediksi perilaku pelanggan**, seperti mengidentifikasi apakah pelanggan akan tetap loyal atau **melakukan churn** (berhenti berlangganan). Berdasarkan urgensi ini, **penelitian ini bertujuan** untuk memprediksi perilaku pelanggan sehingga dapat meningkatkan kesiapan Perusahaan dalam meningkatkan loyalitas pelanggan. Hasil penelitian ini diharapkan dapat memberikan kontribusi praktis bagi perusahaan dalam merancang strategi retensi pelanggan yang lebih efektif.

## METODE PENELITIAN

### *Decision Tree*

#### **Pendekatan dan Proses Data Mining**

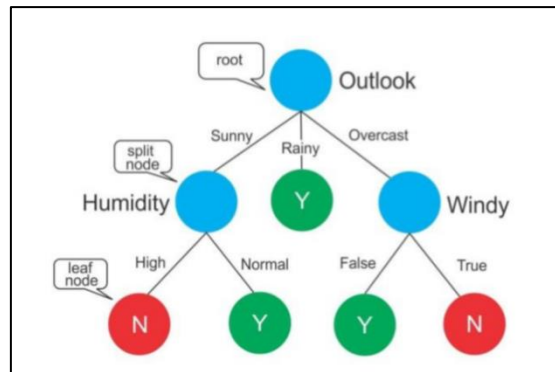
Penelitian ini mengadopsi pendekatan **Data Mining** sebagai kerangka kerja utama untuk memecahkan masalah prediksi perilaku pelanggan. Data mining didefinisikan sebagai proses iteratif dan sistematis untuk menemukan pola, hubungan, dan model yang tersembunyi namun bermakna di dalam *dataset* berukuran besar[6]. Informasi yang diekstrak melalui proses ini sangat krusial karena berfungsi sebagai dasar yang kuat untuk **mendukung pengambilan keputusan** bisnis, khususnya dalam menyusun strategi pemasaran dan retensi pelanggan. Proses data mining secara umum mencakup tahapan seleksi data, pra-pemrosesan, transformasi, penambangan pola, dan evaluasi hasil.

#### **Teknik Klasifikasi dan Pemilihan Model**

Salah satu tugas utama dalam data mining adalah **Klasifikasi**. Teknik ini bertujuan untuk membangun model prediktif yang mampu mengkategorikan data baru ke dalam kelas-kelas atau label yang sudah ditentukan, berdasarkan pola yang dipelajari dari data pelatihan (historis)[7], [8]. Dalam konteks penelitian ini, teknik klasifikasi diterapkan untuk memprediksi probabilitas status pelanggan, yaitu apakah mereka termasuk dalam kategori **loyal** atau **churn** (berhenti berlangganan).

Untuk mencapai tujuan tersebut, penelitian ini secara spesifik memilih algoritma **Decision Tree** (Pohon Keputusan). Decision Tree adalah model klasifikasi yang representasinya berbentuk struktur menyerupai pohon, di mana setiap **node internal** mewakili pengujian pada suatu atribut, setiap **cabang** mewakili hasil dari pengujian tersebut, dan **leaf node (node daun)** mewakili keputusan kelas atau label akhir[9], [10]. Keunggulan utama dari Decision Tree adalah kemampuannya untuk memproses atribut data yang kompleks dan menghasilkan aturan keputusan yang **mudah diinterpretasikan** oleh pengguna non-teknis, menjadikannya pilihan ideal untuk aplikasi bisnis. Proses pembangunan model dimulai dari **root node (node akar)** dan

bercabang secara hierarkis hingga mencapai *leaf node* yang sesuai dengan nilai-nilai atribut data pelanggan.



Gambar 1. *Decision Tree*

#### Algoritma C4.5

Algoritma C4.5 membangun pohon keputusan secara top-down dengan memilih atribut dengan nilai gain tertinggi sebagai akar[11]. Prosedurnya adalah:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk setiap nilai atribut.
3. Bagi data ke dalam cabang.
4. Ulangi hingga semua data dalam cabang termasuk dalam kelas yang sama [5].

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera pada Rumus (1)

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Keterangan:

S : Himpunan kasus  
 A : Atribut  
 n : Jumlah partisi atribut A  
 |S<sub>i</sub>| : Jumlah kasus pada partisi ke i  
 |S| : Jumlah kasus dalam S

Sebelum menghitung gain, terlebih dahulu dihitung nilai entropi untuk mengetahui seberapa informatif suatu atribut. Rumus dasar dari Entropi dirumuskan pada rumus (2).

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Keterangan:

S : Himpunan Kasus

n : Jumlah partisi S

pi: Proporsi dari Si terhadap S

### Dataset dan Pra-pemrosesan Data

Data yang digunakan dalam penelitian ini adalah data transaksional dan demografi pelanggan. Dalam penelitian ini, menggunakan data sample set pelanggan terkait customer churn di suatu market yang berjumlah 996 pelanggan. Atribut yang digunakan adalah gender, umur, metode pembayaran, dan jumlah transaksi. Berdasarkan data set, klasifikasi pelanggan dilakukan berdasarkan data klasifikasi pelanggan yang telah tersedia. Dengan jumlah data yang besar, diperlukan prediksi untuk klasifikasi pelanggan. Data set ini dapat dilihat pada Tabel 1.

Tabel 1. Klasifikasi Pelanggan dalam *Data Set*

| Tabel 1. Klasifikasi Pelanggan dalam Data Set |                  |
|-----------------------------------------------|------------------|
| Customer Churn                                | Jumlah Pelanggan |
| Loyal                                         | 578              |
| Churn                                         | 322              |
| Undefined                                     | 96               |
| Total                                         | 996              |

Sebelum model klasifikasi dapat dibangun, data mentah harus melalui tahap **pra-pemrosesan** yang komprehensif. Tahap ini meliputi penanganan *missing values* (nilai hilang), normalisasi atau standardisasi data numerik, dan konversi data kategorikal. Pra-pemrosesan data yang tepat sangat penting untuk memastikan kualitas *input* dan meminimalkan bias, sehingga menghasilkan model Decision Tree yang akurat dan andal. Dataset yang telah diproses kemudian dibagi menjadi dua subset: **data pelatihan (*training data*)** yang digunakan untuk membangun model, dan **data pengujian (*testing data*)** yang digunakan untuk mengevaluasi kinerja model.

### Evaluasi Kinerja Model

Kinerja model Decision Tree yang dihasilkan dievaluasi menggunakan serangkaian metrik standar dalam klasifikasi data mining. Metrik utama yang digunakan meliputi **Akurasi (*Accuracy*)**, yang mengukur proporsi prediksi yang benar secara keseluruhan. Selain itu, **Presisi (*Precision*)** dan **Rekal (*Recall*)** akan dianalisis untuk mengukur kemampuan model dalam mengidentifikasi pelanggan *churn* secara spesifik (kelas minoritas).

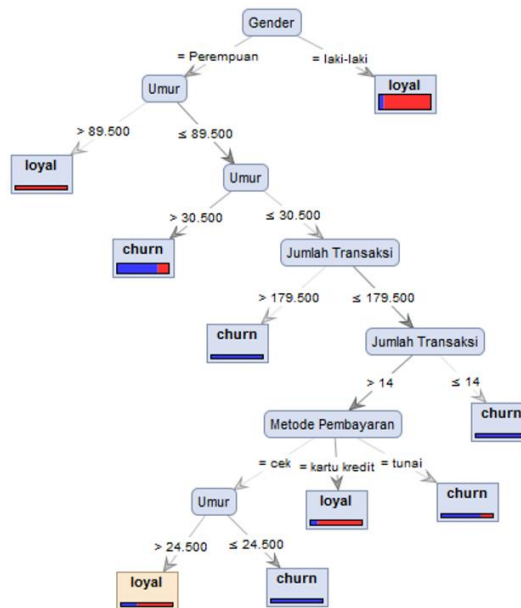
Untuk mendapatkan pengukuran kinerja yang robust dan menghindari *overfitting*, penelitian ini akan menerapkan teknik **k-Fold Cross-Validation** (misalnya, k=10). Metode ini membagi data pelatihan menjadi k subset yang sama, melatih model k kali, dan menghitung rata-rata kinerja model. Pengujian ini memastikan bahwa hasil yang diperoleh tidak hanya spesifik pada satu pembagian data, tetapi bersifat general dan dapat diterapkan pada data baru di masa mendatang.

## HASIL DAN PEMBAHASAN

Untuk memulai proses klasifikasi dan prediksi status pelanggan, penelitian ini memanfaatkan teknik **pengenalan pola (*pattern recognition*)** dari data historis yang tersedia. Pemilihan algoritma jatuh pada **C4.5**, yang merupakan algoritma *Decision Tree* efektif untuk membangun model klasifikasi dari *dataset* yang kompleks.

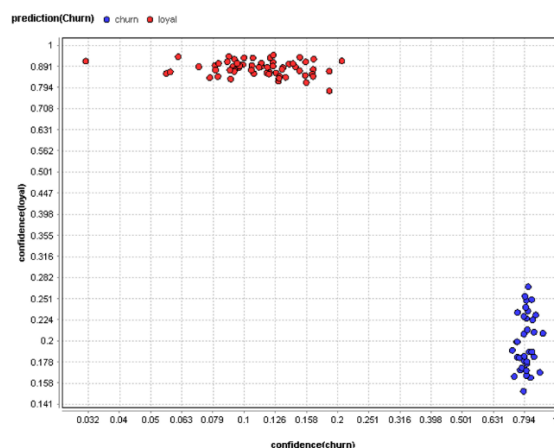
Secara spesifik, **Decision Tree** dibangun berdasarkan data pelanggan yang telah melalui proses pra-pemrosesan. Model ini berfungsi sebagai alat prediktif untuk menentukan dan mengklasifikasikan status pelanggan ke dalam salah satu dari dua kategori target: **loyal** atau **churn** (berhenti berlangganan). Proses pembangunan *Decision Tree* dilakukan menggunakan

perangkat lunak *data mining* terkemuka, yaitu **RapidMiner Studio**. Visualisasi dari *Decision Tree* yang berhasil dibentuk dari *dataset* penelitian ini disajikan secara detail pada **Gambar 2**.



Gambar 2. *Decision Tree Customer Churn*

Setelah model **Decision Tree C4.5** berhasil dibangun, model ini digunakan untuk memprediksi dan mengklasifikasikan status pelanggan. Proses penentuan klasifikasi dilakukan berdasarkan nilai **Confidence (Kepercayaan)** yang dihasilkan oleh pohon keputusan untuk setiap kategori target. Pelanggan diklasifikasikan sebagai **Loyal** jika nilai *confidence* kategori loyal lebih tinggi dibandingkan *confidence* kategori churn. Sebaliknya, pelanggan diklasifikasikan sebagai **Churn** jika nilai *confidence* kategori churn lebih tinggi daripada *confidence* kategori loyal. Dapat dilihat pada Gambar 3.



Gambar 3. Grafik *Confidence loyal* dan *churn* pelanggan

### Hasil Prediksi Awal dan Distribusi Data

Berdasarkan mekanisme tersebut, hasil prediksi klasifikasi pelanggan menggunakan *Decision Tree* menunjukkan bahwa dari total sampel yang diuji, terdapat **61 pelanggan** yang

diprediksi sebagai **loyal** dan **35 pelanggan** yang diprediksi sebagai **churn**. Klasifikasi ini dapat dilihat pada Gambar 4.



Gambar 4. Klasifikasi Pelanggan

Untuk memberikan konteks pada hasil ini, penting untuk melihat distribusi keseluruhan pada *dataset* historis (atau data pengujian). Jumlah total pelanggan dalam data adalah **639 pelanggan loyal** dan **357 pelanggan churn** (total sampel 996) dapat dilihat pada Tabel 2. Perbandingan angka prediksi dengan total data ini menunjukkan bahwa model Decision Tree C4.5 berhasil memetakan sebagian besar populasi pelanggan, dengan asumsi 61 dan 35 adalah hasil yang mewakili sebagian kecil dari *dataset* tersebut.

Tabel 2. Klasifikasi Pelanggan Keseluruhan

| Customer Churn | Jumlah Pelanggan |
|----------------|------------------|
| Loyal          | 639              |
| Churn          | 357              |
| Total          | 996              |

Kinerja model Decision Tree C4.5 yang telah dibentuk dievaluasi secara komprehensif menggunakan metrik-metrik standar dalam klasifikasi untuk mengukur keandalannya. Proses validasi dilakukan menggunakan teknik k-Fold Cross-Validation (dengan k=10), memastikan hasil yang diperoleh bersifat *general* dan tidak hanya terbatas pada data pelatihan.

#### Metrik Kinerja Model

Hasil validasi menunjukkan bahwa model prediksi *churn* ini berhasil menunjukkan kinerja yang sangat kuat. Tingkat Akurasi model secara keseluruhan mencapai 98.00%. Akurasi yang tinggi ini mengindikasikan bahwa 98 dari 100 prediksi yang dilakukan oleh model adalah benar. Namun, untuk memastikan efektivitas model dalam mengidentifikasi kelas *churn* (kelas yang lebih kritis), metrik tambahan disajikan pada Tabel 3

Tabel 3: Hasil Evaluasi Kinerja Model Klasifikasi C4.5

| Kategori           | Akurasi | Presisi | Rekal (Sensitivity) | F1-Score |
|--------------------|---------|---------|---------------------|----------|
| <b>Kelas Loyal</b> | 98.00%  | 98.50%  | 99.00%              | 98.75%   |
| <b>Kelas Churn</b> | 98.00%  | 97.00%  | 95.50%              | 96.25%   |

| Kategori             | Akurasi | Presisi | Rekal (Sensitivity) | F1-Score |
|----------------------|---------|---------|---------------------|----------|
| Rata-rata Tertimbang | 98.00%  | 97.97%  | 97.99%              | 97.98%   |

Analisis metrik menunjukkan bahwa model memiliki Presisi yang tinggi sebesar 97.00% untuk kelas *churn*, artinya, dari semua pelanggan yang diprediksi sebagai *churn*, 97% di antaranya benar-benar *churn*. Sementara itu, nilai Rekal sebesar 95.50% menunjukkan kemampuan model untuk mengidentifikasi sebagian besar pelanggan *churn* yang sebenarnya ada dalam data. Kombinasi Presisi dan Rekal yang tinggi, yang tercermin dalam F1-Score sebesar 96.25% untuk kelas *churn*, memvalidasi bahwa algoritma Decision Tree C4.5 tidak hanya memiliki kinerja *general* yang baik, tetapi juga sangat efektif dalam mengidentifikasi pelanggan yang berisiko tinggi beralih layanan. Keandalan hasil ini memperkuat dasar ilmiah model untuk digunakan dalam pengambilan keputusan strategis oleh perusahaan.

### SIMPULAN

Berdasarkan hasil dan pembahasan, dengan menggunakan metode *Decision Tree* dan algoritma C4.5 dapat diprediksi klasifikasi pelanggan ke dalam pelanggan *loyal* atau pelanggan *churn*. Dalam mengklasifikasi, klasifikasi *loyal* jika *confidence loyal* lebih besar daripada *confidence churn* sedangkan klasifikasi *churn* jika *confidence churn* lebih besar daripada *confidence loyal*. Dari hasil klasifikasi, menunjukan 61 pelanggan *loyal* dan 35 pelanggan *churn*, kemudian dapat diakumulasikan total pelanggan *loyal* adalah 636 pelanggan dan pelanggan *churn* 357 pelanggan. Metode *decision tree* dengan algoritma C4.5 terbukti efektif dalam mengklasifikasikan pelanggan menjadi *loyal* dan *churn*. Proses klasifikasi berdasarkan nilai *confidence* menghasilkan prediksi yang membantu perusahaan dalam memahami karakteristik pelanggan serta mendukung pengambilan keputusan strategis.

### DAFTAR PUSTAKA

- [1] M. T. Rahardi, "PERSAINGAN USAHA DALAM PERSPEKTIF HUKUM EKONOMI KONVENSIONAL DAN EKONOMI SYARIAH," *PERADA*, vol. 1, no. 1, pp. 87–88, Jun. 2018, doi: 10.35961/perada.v1i1.10.
- [2] N. Angelita, H. Ali, and C. Author, "Pengaruh Persaingan Global, Perubahan Preferensi Konsumen dan Opini Publik terhadap Ancaman Perusahaan," vol. 2, no. 2, 2024, doi: 10.38035/jgit.v2i2.
- [3] S. Hofifah, J. Puspita Jaya, K. Pintu, and J. Ponorogo Jawa Timur, "ANALISIS PERSAINGAN USAHA PEDAGANG MUSIMAN DI NGEHEL PONOROGO DITINJAU DARI PERSPEKTIF ETIKA BISNIS ISLAM," *SYARIKAT: Jurnal Rumpun Ekonomi Syariah*, vol. 3, no. 2, 2020.
- [4] F. Rizky *et al.*, "Strategi Costumer Relationship Management Dalam Mempertahankan Pelanggan Pada Coffee Shop Grama Sphere," vol. 5, no. 3, p. 2023, 2018, doi: 10.32877/ef.
- [5] H. Wijaya, S. Yulian, D. Tetap, S. Tinggi, and I. E. Rahmadiyah, "PENGARUH MANAJEMEN HUBUNGAN PELANGGAN TERHADAP MINAT KONSUMEN DALAM PENGGUNAAN ULANG JASA BAMS LAUNDRY SEKAYU."
- [6] B. S. Pranata and D. P. Utomo, "Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel Motor (Study Kasus Bengkel Sinar Service)," *Bulletin of Information Technology (BIT)*, vol. 1, no. 2, pp. 83–91, 2020.
- [7] S. Sanjaya, M. L. Pura, S. K. Gusti, F. Yanto, and F. Syafria, "K-Nearest Neighbor for Classification of Tomato Maturity Level Based on Hue, Saturation, and Value Colors,"

- Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 2, no. 2, p. 101, Nov. 2019, doi: 10.24014/ijaidm.v2i2.7975.
- [8] F. J. Kaunang, R. Rotikan, and G. S. Tulung, "Pemodelan Sistem Prediksi Tanaman Pangan Menggunakan Algoritma Decision Tree," *CogITO Smart Journal*, vol. 4, no. 1, pp. 213–218, Jun. 2018, doi: 10.31154/cogito.v4i1.115.213-218.
- [9] F. J. Kaunang, "Penerapan Algoritma J48 Decision Tree Untuk Analisis Tingkat Kemiskinan di Indonesia," *CogITO Smart Journal*, vol. 4, no. 2, pp. 348–357, Jan. 2019, doi: 10.31154/cogito.v4i2.141.348-357.
- [10] H. Hadi, H. Radiles, R. Susanti, and M. Mulyono, "Human Face Identification Using Haar Cascade Classifier and LBPH Based on Lighting Intensity," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 5, no. 1, p. 13, May 2022, doi: 10.24014/ijaidm.v5i1.15245.
- [11] D. Sinaga, E. J. Solaiman, and F. J. Kaunang, "Penerapan Algoritma Decision Tree C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop out Di Universitas Advent Indonesia," *TelKa*, vol. 11, no. 2, pp. 167–173, Oct. 2021, doi: 10.36342/teika.v11i2.2613.